DOCUMENT RESUME

ED 212 643 TM 810 892

AUTHOR Livingston, Samuel A.
TITLE Estimation of the Conditional Standard Error of Measurement for Stratified Tests.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-81-30
PUB DATE Sep 81
NOTE 10p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Error of Measurement; *Estimation (Mathematics); Mathematical Formulas; *Pass Fail Grading; Scores; *Scoring Formulas
IDENTIFIERS *Stratification; *Test Length

ABSTRACT

The standard error of measurement (SEM) is a measure of the inconsistency in the scores of a particular group of test-takers. It is largest for test-takers with scores ranging in the 50 percent correct bracket; with nearly perfect scores, it is smaller. On tests used to make pass/fail decisions, the test-takers' scores tend to cluster in the range of 80-90 percent correct, with the passing score in the range of 60-70 correct. In this case, the SEM for the full group of test-takers will be much smaller than the SEM for those with scores near the passing score. But, the test-takers with scores near the passing score are the ones for whom the reliability of the test is the most important. For them, measurement errors can make the difference between passing and failing. For this reason, the important SEM is not the SEM for the full group of test-takers, rather, the SEM at the passing score, which will often be substantially larger. A formula for this, and its derivation, are provided. (Author/CE)

RR-81-30

RESEARCH REPORT

ESTIMATION OF THE CONDITIONAL STANDARD ERROR OF MEASUREMENT FOR STRATIFIED TESTS

Samuel A. Livingston

September 1981

Educational Testing Service
Princeton, New Jersey

ABSTRACT

For tests used to make pass/fail decisions, the relevant standard error of measurement is the SEM at the passing score. If the test is highly stratified, this SEM should be estimated by a split-halves approach. A formula and its derivation are provided.

ACKNOWLEDGMENT

I thank Frederic M. Lord for providing an important correction to an earlier draft of this paper.

Estimation of the Conditional Standard Error of Measurement for Stratified Tests

The standard error of measurement (SEM) is a measure of the inconsistency in the scores of a particular group of test-takers. It is largest for test-takers with scores in the range of 50 percent correct; it is much smaller for test-takers with nearly perfect scores. On many tests used to make pass/fail decisions, the test-takers' scores tend to cluster in the range of 80 to 90 percent correct, while the passing score tends to be in the lower tail of the distribution, in the range of 60 to 70 percent correct. In this case, the SEM for the full group of test-takers will be much smaller than the SEM for those test-takers with scores near the passing score. But the test-takers with scores near the passing score are the ones for whom the reliability of the test is most important. For them, error of measurement can make the difference between passing and failing. Therefore, when a test is used to make pass/fail decisions, the important SEM is not the SEM for the full group of test-takers, but the SEM at the passing score, which will often be substantially larger.

One simple solution to this problem is to estimate the SEM at the passing score by the formula for the standard deviation of a binomial distribution:

$$\text{SEM}_P = \sqrt{P(M - P)/M}$$

where P is the passing score (number of correct answers) and M is the maximum possible score, i.e., the number of questions on the test (see Lord, 1957). This solution considers the SEM as the standard error of the sum of a simple random sample of items. For many tests this assumption may be reasonably close to the truth. But on many tests the content is highly stratified and the

test-takers' knowledge varies considerably from one content category to another. In these cases, the binomial formula will tend to overestimate the SEM at the passing score.

Why does the binomial formula overestimate the conditional SEM for stratified tests? It assumes that a test-taker's probability of answering correctly is the same for all questions on the test. But if the test content is highly stratified, this assumption is likely to be quite wrong for many test-takers. For example, ~~support~~ suppose a test contains ten questions, two from each of five content areas. And suppose a test taker knows the right answers to 90 per cent of the questions in the first content area, 30 percent in the second, 70 percent in the third, 20 percent in the fourth, and 90 percent in the fifth. The SEM for this test-taker is actually given by

$$\sqrt{2(.90)(.10) + 2(.30)(.70) + 2(.70)(.30) + 2(.20)(.80) + 2(.90)(.10)}$$

$$= 1.23$$

but the binomial approach yields the estimate

$$\sqrt{10\ (.60)(.40)} \qquad = 1.55$$

because the test-taker's expected total score is six questions, or 60 percent correct.

The usual procedure for estimating the overall SEM (i.e., for the full group of test-takers) on highly stratified tests is to use the split-halves method. The purpose of this paper is to propose an adaptation of the split-halves method for estimating the SEM at the passing score. This solution uses the data from only those test-takers with test scores at the passing score. To

estimate the SEM for these test-takers by the split-halves method, first split the test into halves as similar in content and difficulty as possible. Then compute separate half-test scores for each test-taker whose total (full test) score is at the passing score. Let X_{1i} and X_{2i} represent the two half-test scores for the i th test-taker, and let n represent the number of test-takers with scores at the passing score. Then estimate the SEM at the passing score by the formula:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{1i} - x_{2i})^2}\,.$$

The derivation of this formula is based on the assumption that, for any individual test-taker, the scores on the two half-tests are independent random variables with the same mean and variance. For any given test-taker, the SEM will be the square root of

$$\begin{aligned} &\mathrm{Var}\,(X_{1i} + X_{2i}) \\ &= \mathrm{Var}\,(X_{1i}) + \mathrm{Var}\,(X_{2i}) + 2\,\mathrm{Cov}\,(X_{1i}, X_{2i}) \\ &= 2\,\mathrm{Var}\,(X_{1i}) \end{aligned}$$

because of the assumptions of independence and equal variance of the half-test scores.

To estimate Var (X_{1i}) from the two half-test scores we can use the formula

$$s^2 = \frac{1}{N-1}\sum_{r=1}^{N}(x_{ri} - \bar{x})^2$$

where N = 2 (for the two half-tests) and $\bar{X}$ is the average of the two half-test scores. Thus our estimate of the variance of X_{1i} is

$$\left[x_{1i} - \tfrac{1}{2}(x_{1i} + x_{2i})\right]^2 + \left[x_{2i} - \tfrac{1}{2}(x_{1i} + x_{2i})\right]^2$$
$$= \left[\tfrac{1}{2}(x_{1i} - x_{2i})\right]^2 + \left[\tfrac{1}{2}(x_{2i} - x_{1i})\right]^2$$
$$= \tfrac{1}{2}(x_{1i} - x_{2i})^2 .$$

This estimate refers to an individual test-taker. It can be averaged over all test-takers with total scores at the passing score, yielding the estimate

$$\widehat{\text{Var}}\,(X_{1i}) = \frac{1}{2n}\sum_{i=1}^{n}(x_{1i} - x_{2i})^2 .$$

Since the variance of the total test score is twice the variance of each half-test score, we have the estimate

$$\widehat{\text{Var}}\,(X_{1i} + X_{2i}) = 2\,\widehat{\text{Var}}\,(X_{1i})$$
$$= \frac{1}{n}\sum_{i=1}^{n}(x_{1i} - x_{2i})^2 .$$

The square root of this quantity will provide an estimate of the SEM at the passing score.

REFERENCE

Lord, F. M. Do tests of the same length have the same standard error of measurement? Educational and Psychological Measurement, 1957, 17, 510-521.